INFORMATION MANAGEMENT

INFORMATION IS POWER

The volume of information 'out there' is astounding — some 500 billion web pages and counting, plus an almost infinite quantity of database items. From information we can acquire knowledge and power, but where do you even begin to look for the information, let alone process it?

For many, life without search engines would be almost impossible today, rendering the world wide web all but useless. What would be the benefit of plugging data and information into the web if no-one knows where to find it?

The search function, so dominated by Google that the name has become a verb, has been the defining feature of Web 1.0. People have come to rely on online searches to find what they want, while businesses and anyone seeking to attract online traffic have spent fortunes on the sometimes dark art of search engine optimisation (SEO).

American giants have dominated the world of search, but Europe has been preparing for the next generation of online information management for many years.

The so-called 'future internet' will find European companies competing on level terms.

As the internet has developed we have seen tremendous growth in the availability of non-textual media. Today searching is a complicated affair, not just about finding a word somewhere in an online document. Do you want to search for videos or in videos? Do you want to compare images, perhaps look for someone's face in a video? Maybe you need a fact, perhaps an arrival time for a local train in a foreign country?

When Europe introduced the idea of the 'information society' and the 'knowledge economy' it also realised that some new technologies would be required to turn the mass of information into useful and valuable knowledge.

For this reason it has championed research into novel information management technologies, from innovative search algorithms to personalisation and workflow management.

One area where Europe has built up a strong research base over more than a decade is in the application of semantics to ICT-based information management. The idea is to structure or label data in a way that makes it not just machine readable, but machine understandable. In other words, systems are built that can appreciate the context of data and its meaning.

Semantics is now going online, where structured and non-structured data is carefully indexed using defined tagging methodologies called ontologies. The semantic web offers a powerful new way to access relevant information directly.

EU research on semantics (and the related field of natural language processing, which is trying to develop ways to automate the conversion of unstructured text and speech into semantic information) has made significant contributions to the web ontology language OWL and several semantic mark-up languages.

Several projects have also developed subject-specific ontologies which will make information and data more searchable and accessible across Europe in fields such as healthcare.

»

INFORMATION MANAGEMENT

In 2010 the 'ICT policy support programme' (ICT PSP) had a portfolio of more than 10 projects that were taking natural language and semantic technologies into large-scale pan-European pilot trials. Giants like Yahoo, Facebook and Reuters have all adopted semantic web standards spearheaded in EU projects and the semantic web is now poised for exponential growth.

MORE THAN SEMANTICS

But information management involves more than the methodical tagging and structuring of data. For a quarter of a century the EU has focused its attentions on how ICT can be used as a tool for better business. And in the world of business, information really is an asset — if you know where and how to use it.

Research began under the Union's Fifth Framework Programme (FP5) to explore knowledge management in business. In particular projects have implemented numerous solutions to boost interoperability between systems (especially cross-border information exchange) and more effective information exchange.

Research projects have also promoted the deployment of enterprise resource planning (ERP) systems, supply chain management and human resource management (HRM) solutions across many business sectors.

It is often difficult to see the impact of this research in terms of commercial products, as many of the projects have developed enabling technologies and contributed to standards, rather than commercial products.

But the impact is definitely being felt: you only have to look at what we can do today that we could not just a few years ago — sophisticated forms of analysis using expanding volumes of data in various formats (structured data, text, images, video, 3D, etc.).

European research is at the heart of the intelligent algorithms and in engineering behind the software components that are increasingly required to churn through data and deliver powerful, competitive knowledge. ■

PROJECTS IN FOCUS

Twenty years of sustained research makes Europe a leader in semantic technologies and standards. As the online world is transformed into a semantic web, Europe continues to develop innovative ways for business and administrations to use data efficiently and productively.

Scientific knowledge has been shared in the same way for centuries. But LiquidPublication, a European research project, advocates replacing papers and peer reviews with a new process inspired by the social web.

'The more papers you produce, the more brownie points you get,' says coordinator Fabio Casati. 'So most of your time is spent writing papers instead of thinking or doing science.'

»

Mr Casati wants to tap into the web — including its ability to speed communication, facilitate data storage, search and retrieval, and foster communities of interest — to replace traditional peer reviews and paper publications with a faster, fairer and more flexible process.

'The idea is that when people write papers, they put them on their web-page quickly, easily and for zero cost,' he explains.

Each scientist publishes a journal which can include experiments, datasets, and even blogs as well as papers. Readers add to the publication, make comments and add context and versioning to the publication. The LiquidPublication consortium is putting its ideas into practice, starting with an open source software platform and its own liquid journal on peer review.

European projects have also taken on the wiki world. Wikis are excellent tools for collecting and sharing vast amounts of knowledge among communities. But though wikis are good for storing and retrieving facts (texts, files), they struggle to deliver aggregated information.

Thanks to the work of several EU-funded projects (SEKT, Active and HALO), we now have something called a 'Semantic mediawiki' (SMW) which helps to make the individual pieces of information and knowledge meaningful to computers.

SMW adds semantic annotations that allow a wiki to function as a collaborative database. It was first released in 2005, and currently has over ten developers, and is in use on hundreds of sites. SMW is being considered for adoption by Wikipedia itself.

Meanwhile, European researchers in the four-year NEON project have developed tools that will make it easier to actually create semantic applications — the powerful programs that identify data not just by their textual content, but also by the information's relevance to users.

The tools they created have been tested by such organisations as the Food and Agriculture Organisation (FAO) and by the pharmaceutical sector in Spain. The latest release of the toolkit, one of the core outcomes of the project, is available for download and testing on the NEON website. ■